A VALIDATION STUDY OF A WRITING SKILLS TEST FOR POLICE RECRUIT APPLICANTS

Shelly J. Stolp, B.S.

Thesis Prepared for the Degree of

MASTER OF SCIENCE

UNIVERSITY OF NORTH TEXAS

December 2002

Approved:

Douglas A. Johnson, Major Professor
Terry R. Halfhill, Committee Member
Jefferson L. Sulzer, Committee Member
Earnest H. Harrell, Chair of the Department of Psychology
C. Neal Tate, Dean of the Robert B. Toulouse School of Graduate Studies

Stolp, Shelly J., A Validation Study of a Writing Skills Test for Police Recruit Applicants. Master of Science (Industrial Psychology), December 2002, 83 pp., 17 tables, references, 29 titles.

This study evaluated the effectiveness of a direct test of higher-order and lower-order writing abilities needed for police report writing. This test was specifically designed to address report writing deficiencies experienced by police in the training academy. Descriptive statistics were examined, and relationships between this test and writing ability dimensions included on a separate, indirect, multiple choice test were investigated. Direct and indirect scores were correlated with training academy performance. Because both tests assessed higher-order and lower-order writing abilities, comparisons were made to determine which type of test was most appropriate for assessing the different types of writing skills. Results indicated that the direct test was a valid predictor of academy performance. Direct methods of measurement were found to be better than indirect methods for assessing higher-order writing skills. For lower-order writing skills, the indirect method appeared to be a better measure than the direct method.

TABLE OF CONTENTS

Page

LIST OF TABLES

INTRODUCTION

Well written reports are essential to the efficient functioning of the criminal justice process. However, many police departments are finding that officers lack the skills needed to write effective incident reports. The New Orleans Police Department (NOPD) is not immune to this problem. In a survey of 26 NOPD Field Training Officers, 35% indicated that poor writing ability was a problems for new officers, and 85% judged the dimension of Knowledge of Reports and Basic Writing skills to be the weakest area of job competence for recent academy graduates (Sulzer, 1986).

In an attempt to address this problem, the City of New Orleans Test Development and Validation Department worked with the NOPD to develop the Writing Exercise. This test measures report writing skills and requires applicants to provide a written response to a job related video. The Writing Exercise was the first direct measure of writing ability to be included in the NOPD Police Recruit selection process; prior assessment was done using a multiple choice format only. The current selection battery now includes two measures of the same report writing abilities: The Writing Exercise, which is a direct measure, and multiple choice items from the Written Test, which is an indirect measure. The Writing Exercise was expected to produce a better assessment of applicants writing ability than the multiple choice test had done in the past. This in turn would raise the level of the recruits writing and mitigate past problems of poorly written police reports. Thus, the purpose of this study was to evaluate how well the Writing Exercise fulfilled these expectations.

This study begins with a description of report writing ability deficiencies and the corresponding need for basic skills testing. Next, a review of cognitive tests used in police selection is presented. In this review, studies of general cognitive ability tests for police selection are presented first, studies of ability tests designed specifically for law enforcement selection are reviewed second, and studies focusing on report writing skills are reviewed last. After this review of cognitive tests, two different methods of measuring writing skills, direct and indirect, are defined and discussed. Finally, this review ends by comparing the validity coefficients of direct and indirect measures of writing ability as reported in studies of police selection.

Importance of Report Writing Ability and Well Written Police Reports

The importance of police report writing skills is well documented in the criminal justice literature (Johnson, 1987; Miller & Pomerenke, 1989; Stanard & Associates, Inc.,1992; Wilson & Hayes, 1984). In fact, a nationwide job analysis conducted by Stanard & Associates, Inc. found report writing to be one of the essential functions of a police officer. Wilson and Hayes reported that, not only is the majority of police officers time spent on some type of writing activity, but the ability to apply basic writing skills is a critical factor in the career potential of entry level officers. According to Johnson, officers spend 20-30% of their time writing reports, and supervisors spend 15-20% of their time reviewing these reports. Given this, it becomes apparent that the skills necessary to write effective police reports are crucial for law enforcement officers. In fact, Miller and Pomerenke (1989) stated, Sometimes it seems that the written report is more important than the substantive action that the officer takes in the field (p.66).

Because police reports are extensively used in the court systems, they must be clear and well written. According to D Aulizio and Sheehan (1992a), one of the most important uses of the police report is in the criminal justice system. Well written reports aid in successful criminal prosecutions (Godwin, 1993; Hess & Wrobleski, 1991), can prevent lawsuits from being filed against police departments (D Aulizio & Sheehan,1992a), and reduce the time officers spend testifying in court (D Aulizio & Sheehan, 1992b). In a survey administered to Assistant District Attorneys, Pettaway (1994) confirmed the fact that specific dimensions of well written reports are important for preparing a case and are essential for successful prosecution.

Many people outside the police and justice departments also have access to police reports. Because of this, reports have been described as an extension of the police department s public relations (Miller & Pomerenke, 1989). D Aulizio and Sheehan (1992b) stated that police departments are in the business of providing only two highly related services: police reports and their image within the community. Miller and Pomerenke explained that well written reports create a favorable impression of the police department by demonstrating police competence and credibility. This in turn can lead to public confidence and trust in the police. Finally, Hess and Wrobleski (1991) stated that the efficiency of a police department is directly related to the quality of reports which are read by the public.

The content of a high quality report includes accurate accounts of evidence, witness statements, and a chronology of events occurring in the incident (Pettaway, 1994). A complete report objectively describes facts about who was involved in the

incident, where and when the incident occurred, what happened during the incident, and why the incident took place. Opinions, assumptions, emotional overtones, and police jargon should be left out of the report (Godwin, 1993; Hess & Wrobleski, 1991). Well written police reports are also mechanically correct, and contain few, if any, spelling, punctuation, grammar, or capitalization mistakes (Hess & Wrobleski). Finally, high quality reports are written in such a way that the reader does not misinterpret the information reported. This includes leaving out unnecessary information, writing each sentence in a clear fashion, and logically documenting the incident in the sequence in which the events occurred (Strobl, 1984).

Deficiencies in Report Writing

The law enforcement community and associated agencies have been complaining about the poor quality of reports for years (D Aulizio & Sheehan, 1992b). According to Miller and Pomerenke (1989), much of the time, officers turn in reports that are hastily written, poorly developed, and badly composed. Reports often include incorrect verb tense, wrong pronoun usage, spelling mistakes, and poorly structured sentences and paragraphs (D Aulizio & Sheehan, 1992a). In his review of the police reports of the Walnut Creek, California Police Department, Johnson (1987) concluded that report problems usually fell into one or more of the following six categories: composition and grammar, writing clarity, critical issues discussions, statement and interviews, organization and continuity, and unclear writing standards and expectations.

One of the biggest reasons for these report problems can be attributed to the fact that many recruits do not possess the necessary writing skills needed to complete a police

report (Wilson & Hayes, 1984; D Aulizio & Sheehan, 1992b). In 1974, a study involving 15 California law enforcement jurisdictions indicated that 19% of academy drop-outs were due to lack of skills in reading and writing (Boehm, Honey & Kohls, 1983). Twenty years later, it appears this trend continues as a 1994 survey found that 20% of new police recruits nation wide were rated as poor writers (Pettaway, 1994). For large cities, this number rose to almost 80%. As D Aulizio & Sheehan (1992a) stated, it is no longer safe to assume that police recruit applicants who have graduated from high school possess even the basic writing skills necessary to write valid police reports.

The writing skill deficiencies being reported by police departments appear to mimic the trend in illiteracy throughout the nation. In 1992, Stanard & Associates, Inc. found that, of 4,000 applicants who took a basic skills test, 25% could not read at a twelfth grade level, and 33% could not perform simple mathematics. Moreover, a survey conducted by the U. S. Department of Education s National Center for Education Statistics (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993) estimated that about 21% of Americans over the age of 16 scored in the lowest level of prose literacy. Twenty-two percent of adults scored in the lowest level of quantitative literacy. Based on these results, the U.S. Department of Education estimated that about 40 million Americans do not posses the basic skills needed to compose a letter disputing an error on a credit card statement.

Need for Basic Skills Testing

Without basic skills, individuals would find it difficult to successfully complete police academy training, read and understand general orders and statutes, or write simple

police reports (Stanard & Associates, Inc., 1992). Thus, it is essential for police departments to utilize selection measures that accurately identify individuals who lack the skills needed for successful academy training (Boehm, Honey, & Kohls, 1983). Unfortunately, most cities entrance exams are composed of multiple choice items that require only a moderate level of reading comprehension to complete (D Aulizio & Sheehan, 1992a). Also, due to the multiple choice format of these exams, writing abilities are not directly assessed, if assessed at all. In a survey of 39 police departments, Pettaway (1994) found only three included a writing test in their entry level police officer selection process. Given this lack of writing ability assessment, coupled with high illiteracy rates and school systems that frequently fail to develop basic writing skills (Wilson & Hayes, 1984), it is not surprising that many police departments experience problematic reports. As a first step in overcoming these report writing deficiencies, tests designed specifically for measuring report writing ability need to be added to the entry-level cognitive testing process that is already in place in many police departments.

Review of Studies of Cognitive Assessment in Law Enforcement Selection

To ensure that police recruits enter the academy with the necessary skills, many police departments employ a cognitive ability test in their entry level police officer selection process. A review of the research indicated that a wide variety of cognitive measures are being used to assess applicants skills. However, there appeared to be little consistency among these measures. Some tests assessed global ability constructs while others assessed only a single trait. Also, some tests were developed specifically for law enforcement while others were not. Due to these differences, a first look at studies on

these various measures can seem somewhat overwhelming and confusing.

In an attempt to organize the validation studies of cognitive measures used in police selection, Aamodt (1997) categorized selection measures based on the test developer and degree of commercial availability. The first category, Publisher Developed General Cognitive Ability Tests, comprises tests that were developed by, and can be directly purchased from, national test publishers. Examples include The Nelson-Denny Reading Test (ND) (© Riverside Publishing Company, Itasca, IL, www.riverpub.com), The Wonderlic Personnel Test™ (Wonderlic, Inc., Libertyville, IL, www.wonderlic.com), and The Wechsler Adult Intelligence Scale® - Revised (The Psychological Corporation, San Antonio, TX, www.psychcorp.com). Although not designed specifically for law enforcement selection, these tests measure many constructs related to police performance. The second category, Nationally Developed Law Enforcement Tests, consists of tests designed specifically for law enforcement agencies. More narrow in focus than tests of general cognitive ability, these tests measure specific cognitive abilities needed for police work. Developed by consultants or trade organizations, tests in this category include The Police Officer Selection Test (POST) (© Stanard & Associates, Inc. Chicago, IL, www.stanard.com), and the Law Enforcement Selection Inventory® (LESI)® (Law Enforcement Services, Inc., Greensboro, NC, www.lesi.com). Aamodt s third category contains tests developed by the Federal Government for use either with the military or general employment testing. The fourth and finally category includes civil service exams developed by municipalities for their own use; these exams are not commercially available.

The following review of the validity of cognitive tests used in law enforcement is based on Aamodt s (1997) suggestion that general cognitive measures be compared to tests of specific abilities. The review begins by reporting validity results for general cognitive measures; global in nature, these tests measure a multitude of cognitive constructs. Next, the validities of tests designed specifically for law enforcement are reported; narrowing the cognitive domain, these tests measure only those abilities needed for police work. Finally, this review ends with a summary of studies that included a specific measure of report writing ability.

General Cognitive Measures

In an early attempt to study the predictors of police performance, McAllister (1970) investigated cognitive differences between recruits in his sample that failed the training academy (n = 11), recruits from past academy classes that also failed training (n = 69), and recruits in his sample that passed the academy (n = 396). Results indicated that individuals who failed training had lower scores on The Otis-Lennon Mental Ability Test (© The Psychological Corporation, San Antonio, TX, www.psychcorp.com).

In a similar study, Spielberger, Spaulding, Jolley, and Ward (1979) researched which measures included in a selection battery significantly discriminated between successful and unsuccessful officers. The Nelson-Denny Reading Test (N-D) (© Riverside Publishing Company, Itasca, IL, www.riverpub.com) was the intellectual ability component of the test battery. The N-D is an objective test consisting of four scales: Vocabulary, Comprehension, Reading Rate, and Total. For this study, successes were defined as officers still employed by the department at the end of a one year probationary

period or officers who no longer worked for the department but were considered re-hireable. Failures included recruits that failed the police academy, discharged officers, and officers who resigned and were not considered re-hirable. Of the 317 recruits who passed the selection battery, 233 were classified as successes, 33 were considered failures, and the remaining 51 were never hired by a police agency.

Results of the Spielberger et al. (1979) study were reported separately for Caucasian males, minority males, and females. Significant differences were found on all four N-D scales for Caucasian males, with successes scoring higher on each scale. Females significantly differed on N-D Total and Comprehension scales, with successes scoring higher. Female successes also scored higher on the Vocabulary scale; however, the difference was not significant. Due to small sample size, tests of significant differences were not computed for the minority males; thus, only descriptive statistics were reported for this group. However, on three of the four N-D scales, the mean for successes was higher than the mean for failures.

Kleiman and Gordon (1986) investigated the relationships between intelligence, police academy performance, and job performance. The study was based on the performance of 132 officers who completed police academy training in a large southern city. Intelligence was measured by the Otis-Lennon Mental Ability Test (© The Psychological Corporation, San Antonio, TX, www.psychcorp.com). Academy performance was reported as an overall academic average of several written examinations given throughout the duration of the training. Job performance measures included a behavior checklist (BCL) of eight performance dimensions and a graphic rating scale

(GRS). Each recruit s job performance was rated by two supervisors.

Results of the study indicated that intelligence scores correlated highly with academy performance (0.70). However, the relationships between intelligence and job performance measures were not as strong (0.15 for the BCL and 0.13 for the GRS). Correlations between training performance and job performance dimensions ratings were also low, ranging from -0.03 to 0.18. Further analysis found that intelligence moderated the relationship between training and job performance. The more intelligent individuals had lower and less varied differences between training scores and job performance ratings than those individuals with lower intelligence scores. Thus, for the more intelligent individuals, job performance was more predictable from their training scores.

Hirsh, Northrop, and Schmidt (1986) investigated the validity generalization of cognitive ability tests used in the law enforcement occupations. Forty studies and 381 validity coefficients were included in the research. Of these 381 coefficients, 138 used training performance criteria and 242 used job performance measures. The average sample size of the training studies was 142, and the average sample size for job performance studies was 92. Cognitive tests were categorized by the abilities they measured; categories included clerical aptitude, memory, psychomotor ability, perceptual speed, quantitative ability, reasoning, spatial/mechanical ability and verbal ability. Test of two or more cognitive abilities were classified as a separate test type as were tests that included non-cognitive measures, such as driving practice and human relations.

Results of the validity generalization study indicated that cognitive tests are more predictive of training performance than of job performance. Analysis of studies using

training performance criteria resulted in the following estimated mean corrected correlations: memory (0.41), quantitative (0.63), reasoning (0.61), spatial/mechanical (0.50), verbal (0.64), and verbal plus reasoning (0.71). According to Hirsch et al. (1986), situational specificity could be rejected using the 75% decision rule for four test types: memory, reasoning, spatial/mechanical, and the verbal plus reasoning composite.

For studies using job performance criteria, corrected validity coefficients of the three test types for which situational specificity could be rejected were: reasoning (0.18), spatial/mechanical (0.19), and composite plus human relations (0.31). The estimated mean true validities of the remaining test types ranged from 0.10 and 0.26, and according to Hirsch et al., these results, although small, were generalizable to a majority of situations (p.339).

Measures Designed for Law Enforcement

In one of the earliest studies to involve measures designed specifically for law enforcement, DuBois and Watson (1950) researched the effectiveness of a test battery used by the St. Louis Police Department to screen applicants for the position of probationary patrolmen. Three separate measures of cognitive ability were included in the battery: a general cognitive ability test initially developed for the classification of Army applicants, The St. Louis Police Aptitude Test (PAT); and a Test of English Expression. Training academy performance, scores on an police related achievement test, and supervisory ratings of job performance were the criteria available for the 129 recruits included in this study.

The general cognitive ability test employed in this study was initially developed for use in classification of Army applicants. It includes verbal, numerical, and block sub-scales. The St. Louis Police Aptitude Test (PAT) is an objective test designed specifically for law enforcement by DuBois and Watson (1950). The measure consists of five sections. Section one measures visual memory and requires applicants to remember names and details presented in crime scenes and auto accidents photographs. Section two assesses spelling ability; words such as alias, felony, and accomplice are included in this measure. Reading comprehension is assessed in the third section; reading selections include a fire report, a robbery report, and rules and regulations concerning making an arrest. Section four measures judgment and general information relating to police work, and section five consists of arithmetic problems. The test of English Expression required applicants to write a short essay describing reasons for wanting to become a police officer; essays were rated on coherence based on a five point scale.

Criteria included both training and job performance measures. Academy performance was assessed by a final academy grade, marksmanship scores, and a final score on an achievement test designed to measure general elements of police work. Job performance, as reported as an overall value, was based on supervisory ratings of the following 11 traits: work attitude; loyalty, interest, and enthusiasm; judgment; report writing; investigative ability; alertness; bearing and demeanor; speech; appearance; contacts with the public; and usefulness to the service.

Results of the DuBois and Watson study (1950) indicated that all three predictors significantly correlated with two of the academy performance measures (academy grade

and achievement test performance). Results of the relationship between predictors and academy performance were broken down by class. For class one (n = 72), the general cognitive ability total score significantly correlated with academy grade (0.54). The individual scale correlations, ranging from 0.37 to 0.54, were also significant. For class two (n = 57), the general ability total score correlated significantly with academy grade (0.50), and the individual scale correlations, ranging from 0.30 to 0.53, were significant for this class as well. Correlations between the general cognitive ability test and the Achievement test were computed across both classes (N = 129). These correlations were also significant; the general cognitive ability total score correlation was 0.47, and the individual scale correlations ranged from 0.23 to 0.56.

Analysis of the Police Aptitude Test (PAT) indicated that this exam was a valid predictor. For academy class one, PAT Total score significantly correlated with academy grade (0.39), and four of the five individual scale correlations, ranging from 0.31 to 0.44, were significant. Only the reading scale did not correlate with training performance. For group two, PAT Total score significantly correlated with academy grade (0.50), and four of the five individual scale correlations, ranging from 0.30 to 0.42, were significant. For this group, the spelling scale did not significantly correlate with academy grade. Correlations between the PAT and the Achievement test were computed across classes (N = 129). These correlations were also significant; the PAT Total score correlation was 0.47, and the PAT individual scale correlations ranged from 0.25 to 0.42.

Finally, the English Expression test also appeared to be a valid performance predictor. The correlations with academy grade for the two classes were 0.23 and 0.30

respectively. A significant correlation (0.26) was also found between the English Expression test and the Achievement test (N =1 29).

To summarize, these results indicated that all three tests were valid predictors of two of the academy performance measures: academy grade and achievement test scores. Unfortunately, because no correlations between the different predictors were reported, it is impossible to tell if the abilities measured in each test differed. As for predicting performance, The Police Attitude Test was almost as good as the general cognitive ability test, while the English Expression test proved to be somewhat more limited. Finally, none of the cognitive tests were able to predict marksmanship or supervisory ratings of job performance.

In a second study using cognitive measures designed specifically for law enforcement, Ford and Kraiger (1993) conducted a predictive validation study of the Multijurisdictional Police Officer Examination (MPOE) (© Educational Testing Service, Princeton, NJ, www.ets.org). The MPOE was developed in 1976 by Educational Testing Service to select entry level police officers. Of the 266 applicants who passed the MPOE, 165 completed the academy and were available for the validation study. Criteria included training academy performance, a measure of negative work related incidents, and supervisory ratings of job performance.

As described by Ford and Kraiger (1993), the MPOE is a paper and pencil test of cognitive abilities that were determined, through a job analysis, to be necessary for police officer performance. The exam was developed to assess 12 abilities: flexibility and closure, serial recall, verbal comprehension, spatial scanning, visualization, semantic

ordering, problem sensitivity, induction, memory for relationships, paired associate memory, memory for ideas, and spatial orientation. However, the high internal consistency of the exam (0.95) suggested that the test assessed one general ability. Thus, only a single overall MPOE score was used in the validation study. There are two forms of the MPOE; test-retest reliability was 0.80 and 0.87 respectively for form one and form two. An alternate form reliability of 0.76 was also reported.

Criteria consisted of three measures of performance. The first measure, the Training Academy Composite score, was a sum of five multiple choice exams of police knowledge given in the academy. The second criterion was a measure of negative work behaviors (i.e., citizen complaints, use of force, car accidents, etc.); information concerning these negative behaviors was taken from officers personnel files, coded, and then summed to produce a File Composite score. The third criterion measure was a supervisory rating on nine performance dimensions: criminal investigation; apprehension and follow-up; basic patrol procedures; traffic maintenance and accident investigation; police communication; response to domestic problems and stressful situations; internal relations; judgement, discretion, and common sense; and professional orientation. Because the dimension ratings were highly inter-correlated (0.72), the researchers summed the ratings to form a composite measure of job performance.

Results of the Ford and Kraiger s (1993) validation study indicated that the Multijurisdictional Police Officer Examination (MPOE) was a valid predictor of all three performance measures. The exam significantly correlated with training performance (0.65) and supervisory ratings of performance (0.23). Also, a significant, negative

correlation was found between the MPOE and negative work related behavior (-0.19). From these results, it appears that the MPOE is a valid predictor of both training and job performance.

In a third study in this category, Gruber (1986) conducted a predictive validity study of the Police Applicant Test (PAT), an exam developed by the Calgary Police Service (CPS) to replace a general ability test the department had been using to screen police recruits. Of the 802 applicants that completed the PAT, 66 graduated from the academy and were available for the predictive validity study. Twelve other applicants passed the PAT and were hired, but resigned before criteria data were collected. Criteria for this study included recruit training performance, supervisory rankings of street performance, and early resignations.

As described by Gruber (1986), the PAT was modeled after the MPOE such that is it more specific to police work than a test of general abilities. The PAT was designed to measure four of eleven skills that were determined, through a job analysis, to be essential for the job performance of an entry level zone constable. These skills include verbal communication, judgement, observation, and learning/recall. There are two equivalent forms of the test, each consisting of 140 multiple choice items and one essay. The internal consistency reliability coefficient for form one was 0.86 and 0.87 for form two. The researcher reported a combined internal consistency of 0.86.

Both academy performance and job performance were used as criteria for this study. Academy performance measures included an overall Recruit Training Rating as well as individual grades in the following courses: Criminal Law, Rules of Evidence,

Bylaws and Provincial Statutes, Report Writing, Traffic Laws and Accident Investigation, Human Relations, and Firearms Qualification Score. For the job performance measure, constables were ranked by immediate supervisors from best to worst performers. Reliability coefficients between the supervisors ratings ranged from 0.34 to 0.55.

Results of the Gruber (1986) validation study indicated that PAT was a useful predictor of training performance, on-the-job performance, and early resignations. After corrections were made for range restrictions and attenuation, the PAT significantly correlated with six training measures: Overall Recruit Rating (0.50); Traffic Laws and Accident Investigation (0.42); Report Writing (0.48); Bylaws and Provincial Statues (0.60); Rules of Evidence (0.55); and Criminal Law (0.57). The PAT did not significantly correlate with Human Relations (0.02) or Firearms Qualifications (0.11). Also, although not significant, the uncorrected correlation of 0.19 between the PAT and supervisory rankings of work performance increased to 0.49 after range restriction and attenuation corrections were made. Finally, the group of 12 applicants who resigned was compared to the 66 applicants included in the validation study. Results indicated that the two groups differed significantly on PAT scores; those who resigned tended to have lower PAT scores than those who did not resigned.

In a recent meta-analysis, Aamodt (1997) investigated the relationship between cognitive ability and police performance. Most studies included in this analysis were recent, published between 1970 and 1996. Results of this meta-analysis were based on 47 samples found in 37 studies. Predictor measures ranged from tests of general cognitive ability to tests developed specifically for police work. The cognitive abilities measured

also varied; some studies reported composite scores of multiple ability dimensions, while others only measured a single cognitive construct. Academy performance was a criterion used in 33 of the studies, field training performance was investigated in seven studies, and patrol performance was a criterion in all 37 studies.

Aamodt (1997) found cognitive ability significantly correlated with academy grades, academy graduation, field training performance, supervisor ratings of job performance, peer ratings of job performance, and discipline problems. After correcting for range restriction and attenuation in both the predictor and criterion, Aamodt reported cognitive ability validity coefficients of 0.81 for academy grades, 0.34 for field training performance, 0.36 for supervisor ratings of job performance, and 0.59 for peer ratings of job performance. Validity coefficients, corrected for range restriction and attenuation in the predictor, were 0.56 for academy graduation and -0.27 for discipline problems. Finally, results indicated that, after correcting for all artifacts, variability in all but one of the validity coefficients was attributed to sampling error and measurement artifacts.

The validity coefficients reported in this study demonstrate the usefulness of cognitive ability tests for law enforcement personnel selection. However, unlike the work of Hirsch et al. (1986), this study did not include a separate analysis for each of the different types of cognitive ability measures. Acknowledging this limitation, Aamodt stated that further research is needed to determine what types of cognitive ability are being measured in each of the different tests that were included in his meta-analysis. In concluding his research article, Aamodt called for research that compares the validity of general cognitive ability tests to the validity of tests of specific abilities.

Measures of Police Report Writing Ability

Unusual among tests of cognitive ability in police selection, The National Police Officer Selection Test (POST) (© Stanard & Associates, Inc. Chicago, IL, www.stanard.com), developed by the consulting firm Stanard & Associates, Inc. (1991), includes a sub-scale designed specifically to measure report writing ability. The POST also measures three other skills that were found to be essential for learning and performing police officer duties. Section I assesses basic arithmetic skills (i.e., addition, subtraction, multiplication, division, and percentage calculations); examples of the multiple choice test items in this section include calculating mileage on a patrol car and overtime hours. Section II measures reading comprehension; applicants are required to answer multiple choice and true/false questions based on readings of laws, court cases, and general orders. Section III assesses spelling, grammar, and punctuation abilities; this section, composed of multiple choice items, requires applicants to identify misspelled words and correctly complete sentences containing police related material. The last section, Incident Report Writing, assesses the ability to write complete sentences with correct spelling, punctuation, and grammar; applicants are required to read an incident report and then provide written answers to questions about the incident. Reliability coefficients of 0.79 and 0.80 were reported for first three sections and the Report Writing section respectively (Rafilson & Sison, 1996).

Numerous validation studies have been conducted on the National Police Officer Selection Test (POST). Stanard & Associates, Inc. (2001) reported the results of a concurrent validation study investigating the use of the POST as a predictor of police

academy performance. One hundred twenty three applicants and recruits took the POST and the SRA Verbal test which, according to Stanard & Associates, Inc., is a well established measure of cognitive ability.

Results indicated that the four POST scales measure different abilities. Inter-correlations between scales ranged from 0.23 to 0.48; the Math and Reading scales had the highest inter-correlation. Correlations between The SRA Verbal score and the POST scores also suggested that the two cognitive tests, although similar, may be taping different abilities. Correlations between the two tests ranged from 0.32 to 0.58; the Report Writing scale had the lowest correlation and the Total POST score had the highest.

Correlations of these various predictors with final academy scores were as follows: POST Math (0.31), POST Reading Comprehension (0.53), POST Grammar (0.42), POST Report Writing (0.40) Total POST (0.55) and SRA Verbal (0.37). These results indicated that, as designed, the POST measured specific, separate abilities. Also, for this study, the POST was better at predicting academy performance than the SRA, which is a test of general cognitive ability.

Rafilson and Sison (1996) reviewed several criterion related validity studies conducted with the POST, three of which included academy performance as criteria. Sample sizes in these studies were 98, 186, and 193. In all three studies, significant correlations were found between academy performance and each of the four POST scales. The ranges of the reported correlation for the POST were as follows: Mathematics (0.44 - 0.47), Reading Comprehension (0.43 - 0.45), Grammar (0.20 - 0.40), and Report Writing (0.24 - 0.42). Two studies reported significant correlations with the POST Total

score (0.54 and 0.58). In another reviewed study, a significant correlation of 0.58 was found between police report evaluations of 38 police officers and their POST Report Writing scores. Finally, another study reported a significant correlation, corrected for attenuation, of 0.32 between POST Total score and supervisory ratings of critical thinking for 246 law enforcement incumbents. According to Rafilson and Sison, these results support the use of the POST as a valid predictor of both training and work performance.

In another study investigating the use of report writing ability measures, Berkley (1997) validated a Corrections Officer Training (COT) entrance exam developed by the Pennsylvania State Civil Service Commission. The COT included two cognitive ability tests customized for law enforcement: The Written Test and The Writing Exercise. The Written Test was a multiple choice exam consisting of four sub-tests: observation, following oral instructions, understanding rules and regulations, and following written instructions. The Writing Exercise, designed to only assess writing ability, required applicants to provide a writing sample. The writing sample was rated on two scales: accuracy and completeness, and clarity of expression. Based on COT exam scores, 506 applicants were admitted into, and completed, the training academy.

Criteria included in this study were academy test scores and job performance ratings. The academy test score was the sum of four exams given after each week of academy training. Job performance ratings consisted of scores on an evaluation form designed to match the skills and abilities measured in the selection tests. These forms were filled out by two Training Sergeants and one Training Lieutenant after recruits completed at least 10 weeks of on-the-job training.

Results of this study suggested that the Written Test and the Writing Exercise are measuring different abilities as the inter-correlation between the two tests was 0.50. Correlations between the four sub-tests of the Written Test indicated that, as intended, each measured a different skill or ability (Berkley, 1997). However, the Writing Exercise appeared to measure only one factor as each scale (accuracy and completeness, and clarity of expression) correlated highly with the total writing scale, 0.93 and 0.84 respectively.

As for predicting academy scores and job performance, results indicated that both tests were valid. For the Written Test total, corrected validity coefficients were 0.43 for job performance and 0.72 for academy performance. No scale validity coefficients were reported. Corrected validity coefficients for the Writing Exercise total and job and training performance were 0.28 and 0.30 respectively. These results indicated that although the Written Test is more highly correlated with performance measures, the Writing Exercise is a valid measure of a separate ability.

Table 1 provides a summary of the results reported in the reviewed studies of cognitive ability used in law enforcement selection. There are several important findings that are worth noting. First, test of general cognitive ability consistently appeared to effectively predict recruits training performance. Second, cognitive ability measures developed specifically for law enforcement were similar to the general cognitive ability measures when used for predicted training success. Third, the relationship between cognitive measures and job performance was less pronounced and less consistent. Finally, the majority of the law enforcement selection tests did not include a measure of report writing ability.

Table 1

Summary of Results Found in the Literature Review of

Cognitive Ability Measures in Law Enforcement

		Results and/or Validity Coefficients	
Study	Test Type	Academy Training	Job Performance
McAllister, (1979)	General	Training failures had lower scores	Not measured
Speilberger et al., (1979)	General	Training successes had higher scores.	Not measured
Kleiman & Gordon, (1986)	General	0.70	0.15 and 0.13
Hirsh et al., (1986)	Meta analysis	0.41 - 0.71	0.10 - 0.31
DuBois & Watson, (1950)	General	0.47 and 0.54	No relationship found
	Police related (PR)	0.39 and 0.50	No relationship found
	Report writing (RW)	0.23 and 0.30	No relationship found
Ford & Kraiger, (1993)	Police related	0.65	0.23 and -0.19[a]
Gruber, (1986)	Police related	0.42 - 0.50	0.49
Aamodt (1997)	Meta-analysis	0.34 - 0.81	-0.27[a] - 0.59
Stanard & Associates, Inc., (2001)	General	0.37	
	PR + RW	0.55	
Rafilson & Sison, (1996)	PR + RW	0.54 - 0.58	0.32 - 0.58
Berkley, (1997)	Police Related	0.72	0.43
	Report Writing	0.30	0.28

Note. Validity coefficients represent correlations with the Total score.

[a] Negative correlation found between test score and negative work behaviors.

This lack of report writing ability assessment may be contributing to report writing problems reported by many police agencies. Of all the selection tests reviewed, only three directly assessed the writing skills necessary for effective report writing: The test of English Expression (DuBois & Watson, 1950), The POST (Stanard & Associates, Inc., 1991), and the Writing Exercise (Berkley, 1997). However, the writing ability constructs measured in these three exams (coherence; grammar, spelling, and punctuation; accuracy and completeness, and clarity of expression) appeared to be dissimilar, and for the most part, not well defined. Test format (multiple choice, short answer, essay) also differed between measures. This lack of consistency in the research suggests that the constructs of report writing have yet to be refined. This is not surprising as the literature review indicated that personnel selection researchers have only recently realized the importance of report writing ability assessment for entry level police officers.

Assessment of Written Communication

Although the assessment of written communication has received very little attention in personnel selection, it has been extensively covered in educational studies. According to the educational research, (Ackerman & Smith, 1988; Breland & Gaynor, 1979; Moss, Cole, & Khampalikit, 1982; Quellmalz, Capell, & Chou, 1982), writing ability measures are typically divided into two categories: direct assessment and indirect assessment. Direct measures require the generation of a writing sample which is then scored by one or more raters. Indirect measures, also called objective assessment, require no writing at all. Instead, students usually read a sentence and then select, from one of four choices, a better alternative to the given sentence (Moss, Cole, & Khampalikit).

The effectiveness of both writing ability measures has been heavily debated within the educational research (Quellmalz, Capell, & Chou, 1982). According to Breland and Gaynor (1979), proponents of direct measures argue that essay tests are better because they require an actual sample of the abilities being measured, have higher face validity, and assess larger elements of composition such as unity, organization, and content. Opponents of indirect measures also state that objective measures are less adequate because they tend to only measure sentence-level mechanics (Quellmalz, Capell, & Chou, 1982) and require only editing and reading skills (Ackerman & Smith, 1988). Supporters for indirect measures argue that multiple choice tests tend to have higher reliability, are more efficient to give and score, and, because items are usually not related, are less affected by a mistake on any one item (Breland & Gaynor).

Quellmalz, Capell, and Chou (1982) suggested that the debate about which measure is more effective may be irrelevant because each measure appears to assess different abilities. After reviewing numerous studies comparing direct and indirect writing measures, the researchers found that scores on the two measures were only moderately correlated, with the majority of correlations ranging from 0.43 to 0.68. Quellmalz et al. suggested that higher correlations were not found because the reviewed studies were based on one of the following assumptions: (1) both sets of measures tapped either a unidimensional writing ability or, (2) the same set of writing skills is being measured by both. According to Quellmalz at al., higher correlations are not found because the tasks required by different response modes (paragraphs, essays, and multiple choice items) provide different information about writing ability.

To test this hypothesis, approximately 200 high school students were given three constructed response tasks (two essays and one paragraph) and a multiple choice test that measured the same dimensions as the direct measures. Paragraphs and essays were scored on five dimensions: (1) General Impression, (2) Focus, (3) Organization, (4) Support, and (5) Mechanics. Quellmalz at al. (1982) defined the dimensions as follows:

> General Impression - a global judgement of writing quality assigned by raters after a quick initial reading of the writing sample.
>
> Focus - the extent to which the subject and main idea of the writing sample were clearly stated or implied.
>
> Organization - the extent to which the main idea was developed according to a discernible method of organization.
>
> Support - the extent to which generalizations and assertions were supported by specific, relevant, subordinate statements.
>
> Mechanics - the extent to which the writing sample was free from intrusive sentence-level mechanical errors (e.g., usage, sentence structure, spelling, capitalization and punctuation). (p.245-246)

The multiple choice test assessed three of these same dimensions: (1) Focus, (2) Organization, and (3) Support.

Data analysis indicated that the General Impression rating was inseparable from Focus and Organization ratings. Thus, this trait was excluded from subsequent analysis as it appeared to contain little or no additional information. Analysis of scores on the remaining traits (Focus, Organization, Support, and Mechanics) suggested that the writing assessments were measuring three factors rather than the proposed five. Focus and Organization loaded on a Coherence factor, while Support and Mechanics remained as the other two factors. Within each factor, the multiple choice trait scores, with the

exception of Organization, loaded lower than the corresponding essay variables. The researchers attributed the Organization loading to the more general nature of the Coherence factor. Also, the low loading of the multiple choice Support trait on the Support factor suggested that direct and indirect measures were not comparable on this trait. Overall, these findings suggested that the information obtained from indirect and direct measures may be similar is some ways and dissimilar in others.

In another review of the relationship between direct and indirect measures of writing ability, Ackerman and Smith (1988) reported findings similar to Quellmalz et al. The Ackerman and Smith review found correlations between the two measures ranged from 0.30 to 0.60. Based on this finding, the authors concluded that direct and indirect measures are not equal in value because they may be assessing dissimilar skills.

To explain possible reasons for the lack of agreement between information provided by direct and indirect measures of writing ability, Ackerman and Smith (1988) expanded on a model of writing process proposed by Hayes and Flower (as cited in Ackerman & Smith). According to this model, the writing process consists of three interdependent components: planning, translating, and reviewing. The planning component is comprised of idea generation, idea organization, and goal setting. The translating process requires generated ideas be translated into grammatically complete sentences. The reviewing component involves improving the quality of written material by checking for mechanical errors. According to the Ackerman and Smith model, direct and indirect assessments differ on these writing process components; direct measures include all three components whereas indirect measures only include the reviewing

component. It was hypothesized that this difference may be attributing to the moderate correlations often reported between direct and indirect measures of writing ability.

To test their model, Ackerman and Smith gave 219 students the Comprehensive Assessment Program (CAP) battery (a multiple choice measure), an essay task, and a free-response version of the CAP. The only difference between this measure and the multiple choice test was that the free response test required students to generate correct answers rather than select correct alternatives. All three tests were scored on the same sub-scales: (1) Spelling, (2) Capitalization/Punctuation, (3) Correct Expression, (4) Usage, (5) Paragraph Development, and (6) Paragraph Structure.

Results of the study supported the hypothesized model which suggested that direct and indirect writing ability measures provide different information on writing ability. Specifically, as predicted, the essay test was a more accurate measure of idea generation than the multiple choice test. In fact, results indicated that essay scores were almost totally dominated by the higher-order generation components such as paragraph development and paragraph structure. The free-response measure was better than the multiple choice test in assessing the ability to organize coherent paragraphs. Results also indicated that multiple choice measures represent editing and reviewing skills. Because each measure appeared to assess different abilities, the authors concluded the study by suggesting that a reliable measure of writing ability includes both direct and indirect measures.

From the review of direct and indirect writing assessments, it appears that these two measures provide both similar and unique information about writing ability. In

general, results suggest that direct measures are better assessments of higher-order generation skills such as idea development, paragraph development, and structure. Indirect measures appear to be measuring lower-order skills such as spelling, capitalization, and punctuation. Given the differences in information obtained from direct and indirect measures of writing ability, high correlations between these two measures should not be expected. Therefore, tests that include both direct and indirect measures appear to be the most reliable predictors of overall writing performance.

Assessment of Written Communication Relative to Police Reports

Writing police reports requires both mechanical and higher-order writing skills. According to the literature, an applicant s spelling and punctuation abilities may be best assessed by a multiple choice test, while an applicant s ability to report the facts of an incident in a logical sequence may be more effectively measured by an essay exam. Moreover, although skills such as spelling and punctuation are necessary for report writing, the literature suggests that the report writing process relies more heavily on higher-order writing abilities such as idea generation and organization. When comparing direct and indirect measures of writing ability, it appears that essay tests may be more effective at assessing some of the more important abilities needed to produce a police report. Furthermore, according to Ackerman and Smith (1988), the most reliable report writing ability measure would include both multiple choice and essay items.

For selection researchers the next, and more important step, is to determine how well both measures correlate with academy training or job performance. Table 2 presents a summary of validity coefficients of report writing ability measures that were found in

Table 2

Comparisons of Validity Coefficients Between Direct and Indirect Writing Ability Tests as Predictors of Academy Performance

Test and Study	*N*	Dimension	Test Type: Direct	Test Type: Indirect
Higher-order Abilities				
Test of English Expression (DuBois & Watson, 1950)	129	Coherence	0.23	
Writing Exercise (Berkley, 1997)	506	Accuracy and Completeness & Clarity of Expression	0.30	
Lower-order Abilities				
POST (Rafilson & Sison, 1996)	38-193	Spelling, Punctuation & Grammar	0.24 - 0.58	0.20 - 0.42

Note. The range of POST validity coefficients represents the results of five studies (*N* = 123, 38, 98, 193, and 186).

the reviewed studies. The summary is broken down by test mode (direct vs. indirect) and trait level (higher-order vs. lower-order). Although only four measures were included in this summary, there are several points worth noting. First, both essay and multiple choice measures appear to be valid predictors of training performance. Second, and somewhat surprising, the POST direct measure appeared to account for more variance in training performance than the POST indirect measure. Perhaps this is due to the complexity of writing tasks required in the academy. Finally, no validity coefficients for a multiple choice test (indirect measure) that measured a higher-order writing ability were reported. Because of this, comparisons could not be made between the predictive validity of direct and indirect tests of higher-order writing skills.

Purpose of this Study

The purpose of this study was to evaluate the effectiveness of the Police Recruit Writing Exercise, a direct test of both higher and lower-order writing abilities. This test was specifically developed to address the report writing deficiencies experienced by recruits in the training academy. Unlike past measures of written communication (i.e., multiple choice Written Test), the Writing Exercise required police recruit applicants to produce a written report describing what they saw in a job-related video. This report was then scored by two trained raters. It was anticipated that this new direct test of writing ability would improve the assessment of applicants report writing skills, and this in turn, would help reduce report writing problems noted by the police academy.

Various psychometric properties of the Writing Exercise were investigated. Reliability was estimated by a coefficients of equivalence and parallel-forms. The

criterion-related validity of the Writing Exercise in predicting recruits writing abilities was estimated by examining its relationship with training academy performance measures. Finally, because applicants report writing ability was also assessed with an indirect multiple choice measure, convergent and discriminate validities were examined by the multitrait-multimethod matrix (Campbell & Fiske, 1959).

Differences between indirect and direct report writing ability tests were also studied. To determine which measure better predicted academy performance, the validity coefficient of the Writing Exercise Total (WET), which is a direct measure of writing ability, was compared to the validity coefficient of the Multiple Choice Total (MCT) which is an indirect measure of writing ability. Given the research reviewed in this area, it was hypothesized that the WET validity coefficient would be greater than the MCT validity coefficient, and the combined measures (MCT+WET) validity coefficient would be greater than the WET validity coefficient alone: MCT < WET < (MCT+WET).

Lastly, because the multiple choice Written Test included a measure designed to assess higher-order writing abilities, the literature gap regarding the validity of an indirect measure of a higher-order writing ability was addressed. Based on the reviewed research, it was hypothesized that the Writing Exercise measure of the higher-order trait would better predict academy performance than the multiple choice measure of the same trait.

METHOD

Participants

Participants for this study were drawn from applicants who took the New Orleans Police Recruit Exam from 1996 to 1999. This exam was a multiple hurdle exam; the first hurdle was the multiple choice Written Test and the second hurdle was the Writing Exercise. Only those applicants that had at least a high school diploma (or equivalency), possessed a valid driver s license, and were 20 years of age or older were eligible to take the multiple choice Written Test. The total sample for the multiple choice Written Test included 3,250 applicants with 1,726 African Americans, 1,159 Caucasians, 106 Hispanics, 27 Asians, and 196 others who did not specify their race. Applicants who achieved a passing score on the multiple choice Written Test qualified to take the Writing Exercise. The sample for the Writing Exercise included 2,670 applicants with 1,402 African Americans, 1,124 Caucasians, 93 Hispanics, 21 Asians, and 30 others who did not specify their race. Of these 2,670 applicants, 625 successfully completed the remaining selection hurdles (background check, psychological screening, and medical screening) and entered one of eleven police academy classes that were conducted during this time interval. Due to resignations and terminations from the academy, criterion data were available for 567 recruits with 316 African Americans, 223 Caucasians, 21 Hispanics, 4 Asians, and 3 others who did not specify their race. Table 3 presents the various sample sizes for the predictor and criterion variables.

Table 3

Samples Sizes of Predictors and Criteria

Measure	*N*
Predictors	
Police Recruit Written Test	3,250
Police Recruit Writing Exercise	2,670
Criteria	
Academy Performance	567

Predictor Measures

Police Recruit Written Test

The Police Recruit Written Test, developed by New Orleans Civil Service, was the first exam in the multiple hurdle Police Recruit selection process. The Written Test is a job relevant multiple choice exam consisting of 215 questions designed to measure 15 knowledge, skills, and abilities (KSAs) that subject matter experts (SMEs) identified as necessary for success on the job. Table 4 presents each KSA measured by the exam and the number of questions devoted to each. Of the 15 KSAs, the following three were identified as measures of police report writing ability:

(1) Written Communication--Ability to communicate in writing what occurred in an incident accurately and clearly.

(2) Form Completion--Ability to learn how to complete forms with some instructions.

(3) Knowledge of Grammar, Spelling, and Punctuation–Knowledge of basic grammar, spelling, and punctuation to properly write and complete reports. This includes the knowledge of when grammar is used correctly and when a narrative is clear and intelligible.

The Written Communication KSA was indirectly assessed by having applicants select, from four alternatives, the sentence that best communicated what was seen in a photograph. The Knowledge of Grammar, Spelling, and Punctuation KSA was indirectly assessed by having applicants read four sentences for grammar, punctuation, and spelling. Applicants then had to select from the four choices the sentence that was the most correct. The Form Completion KSA was indirectly assessed by having applicants review a section

Table 4

Knowledge, Skill, & Ability Dimensions Measured in the Police Recruit Written Test and the Number of Test Items Measuring Each Dimension

Dimension	Number of Items
Perceptual Accuracy	65
Verbal Memory	9
Associative Memory	8
Selective Attention	8
Oral Comprehension	14
Observation Ability	20
Written Communication	8
Deductive Reasoning	12
Inductive Reasoning	11
Reading Comprehension	9
Exercising Judgement	20
Form Completion	8
Recognizing Transformed Objects	5
Grammar, Spelling, Punctuation	6
Social Judgement/Interpersonal	12

of an Incident Report along with a corresponding form of the incident. Applicants had to choose, from four choices, the part of the form that was filled out incorrectly. Appendix A provides sample items from each of these three subsections. The score on each section was the total of questions answered correctly.

Police Recruit Writing Exercise

The Writing Exercise was developed in 1996 and is an extension of the New Orleans Police Department Communication Test (Pettaway, 1994), a pilot test developed and validated in 1994. The Writing Exercise consisted of two job-related scenarios presented on video that depict officers responding to a police incident and collecting information from the individuals involved. The first video involved officers responding to a domestic disturbance call and questioning the victim of the assault. The second video involved officers responding to an auto theft. While watching the videos, applicants were instructed to take detailed notes. After the end of each video, applicants were given 20 minutes to write a narrative report on the incident and were instructed to use their notes while writing the narrative.

Each report was rated on two scales: (1) Content and Completeness, and (2) Writing Mechanics. Each scale assessed report writing abilities that were determined to be important for the position of Police Recruit. These abilities were derived from job analysis, police reports, police training academy courses, report writing surveys, and SMEs from the New Orleans Police Department and District Attorney s Office.

The Content and Completeness Scale was designed to measure the Written Communication KSA. This scale was defined as the degree to which a report accurately

and completely includes the details of the incident and avoids leaving the reader with unanswered questions about the incident. This scale included the following two sub-scales:

(1) Accuracy in Description--The five necessary elements (who, what, where, when, why) are accurately presented in moderate detail.

(2) Accuracy in Statement Content--Report contains accurate information regarding witness statements, includes reporting everything the witnesses stated.

The Writing Mechanics Scale was designed to assess the Knowledge of Grammar, Spelling, and Punctuation KSA. This scale was defined as the degree to which the report clearly conveys information about the incident. This includes presenting information in a organized fashion; properly structuring sentences; and correctly using grammar, verb tense, spelling, and punctuation. The following six sub-scales were included in this scale:

(1) Logical Flow--Report is arranged chronologically or in some other logical manner.

(2) Vocabulary--Words in the report are correctly used and the report contains no slang.

(3) Spelling and Capitalization--Words in the report are spelled and capitalized properly.

(4) Punctuation--Report contains correct punctuation which includes commas, colons, apostrophes, semicolons, quotation marks, and ending punctuation.

(5) Grammar--Report contains correct grammar which includes correct subject/verb agreement and verb tense, no vague pronouns, and complete sentences.

(6) Logical Sentence Structure--Individual sentences in the report are structured correctly. Sentences include parallelism and internal coherence and contain no misplaced modifiers, mixed structures, or run-on sentences.

Likert-type rating scales that ranged from 1 (*poor*) to 5 (*excellent*) were used to assess applicants performance on each sub-scale. Behavioral anchors were given for ratings 1, 3, and 5. A team of two raters, using the rating scales, rated both of the applicant s written reports. Each rater first independently rated the applicants report and then compared their rating with their partner s. Rating differences greater than one point were discussed, and when applicable, initial ratings were modified. Raw scores were summed across rater and exercise, totaling four ratings for each sub-scale. These raw scores were then standardized and weighted. Weights were determined by subject matter experts who were asked to ascribe a percentage to each sub-scale to show its importance relative to the other sub-scales. Weighted sub-scales scores were then summed to produce two scale scores and a total score. Table 5 lists the scales, sub-scales, weights, and the KSA measured by each scale.

Writing Exercise raters had to have at least an undergraduate degree and attend a mandatory half day training session on the rating process. The training session included an extensive review of the scales, a discussion on how to avoid the most common rating errors, instructions on how to properly fill out rating forms, and procedures for rating with a partner. To ensure that raters understood the rating scales and the rating process, the writing samples of hypothetical candidates were rated, and these ratings were discussed.

Table 5

Writing Exercise Scales and Sub-scales with Corresponding Weights and KSAs

Scales & Sub-scales	Weight	KSA Measured
I. Content and Completeness Scale		Written Communication
Accuracy in Description	0.18	
Accuracy in Statement Content	0.23	
II. Writing Mechanics Scale		Knowledge of Grammar, Spelling, and Punctuation
Logical Flow	0.15	
Vocabulary	0.10	
Spelling and Capitalization	0.06	
Punctuation	0.06	
Grammar	0.10	
Logical Sentence Structure	0.11	

Criterion Measures

Police training academy performance measures were used as the criteria in this validation study. The police academy standards match the standards set forth by Peace Officers Standards and Training (POST). This State Agency mandates the minimum standard of training for all Louisiana Peace Officers. These standards are fixed by State Law and cannot be waived by an individual or department. The New Orleans Police Department developed its recruit training curriculum to match these standards. The criteria used in this validation study were academy graduation, overall academy performance, and performance in each of the academy s standardized courses listed below:

(1) Legal Aspects--This course covered the history of law enforcement and oriented recruits to the criminal justice system at federal and state levels. Laws relevant to police work were reviewed; these include, but were not limited to, Probable Cause, the Exclusionary Rule, Arrest and Entrapment, Confession an Admissions, Vehicle Stops and Seizures, and Municipal Criminal Code. This class was academic, requiring reading and memorization. Assessment consisted of multiple choice and true/false exams.

2) First Aid--This course covered first aid techniques and cardio-pulmonary resuscitation. Standards of performance were based on those set by the Red Cross. Less academic in nature, this course required demonstration of competency in first aid techniques. Assessment consisted of multiple choice test items and ratings of performance in practicums.

(3) Investigations--This course explained steps of proper investigations from initial interviews and interrogations to evidence and fingerprinting. Investigation techniques were reviewed specific to particular crimes (i.e., theft, burglary, homicide, sex crimes, drug crimes). The class was academic in nature, requiring a large amount of reading. Assessment consisted of multiple choice test items.

(4) Report Writing--The course covered contents of police reports, departmental standards for report writing, and procedures for filling out various forms. Recruits had to produce incident reports based on provided, hypothetical information. Writing skills were assessed with short answer test items and ratings of incident reports. The incidents reports were assessed on proper form completion; accuracy and completeness of reported incident; and correct use of grammar, spelling, punctuation, and vocabulary.

(5) Traffic--This course reviewed the basics of accident investigations; state and municipal motor vehicle laws were explained. Requirements included writing traffic reports on several different accident scenarios; this included filling in forms and writing a narrative of the incident. Reports were assessed on proper form completion, accuracy and completeness of the reported incident, and correct use of grammar, spelling, punctuation, and vocabulary.

(6) Patrol Activities--This course reviewed the knowledge and skills necessary for police patrol. Topics included the basic methods of patrol as well as methods specific to particular incidents such as vehicle stops tactics, violent crimes, and unusual occurrences. Radio procedures and radio user training were included.

This class was more practical than academic, requiring demonstration of competency in patrol activities. Assessment consisted of multiple choice test items and ratings of performance in practicums.

(7) Officer Survival--This was a practical skills course covering the basics of officer survival. Recruits were trained on skills such as baton handling, handcuffing techniques, weapon retention, and crowd control. This course was more practical than academic, requiring demonstration of competency in officer survival skills. Assessment consisted of ratings of performance in practicums.

(8) Specialized Activities--This course taught the proper procedures to follow for specific incidents such as domestic violence, auto theft, the handling of juveniles, hostage negotiations, and the handling of the mentally disturbed. Police personnel with expertise in the topic to be covered provided the lecture, emphasizing job-related experience. Assessment included multiple choice and true/false test items.

(9) Police Community Relations--Topics covered in this course included police ethics, minority relations, building respect for the police, and crime prevention education. Police personnel with expertise in the topic to be covered provided the lecture, emphasizing job-related experience. Assessment included multiple choice and true/false test items.

Variables

Indirect Writing Ability Scores

Three dimension scores from the Written Exercise provided the data for the indirect measure of writing ability. These scores were also summed to produce an

indirect, multiple choice composite score. Thus, the four indirect predictor scores included in this study were: (1) Written Communication (2) Form Completion (3) Knowledge of Grammar, Spelling, and Punctuation and (4) Multiple Choice Total. For the purposes of this study, the Written Communication was classified as a higher-order writing ability and Knowledge of Grammar, Spelling and Punctuation was classified as a lower-order writing ability.

Direct Writing Ability Scores

All eight ability ratings from the Writing Exercise provided the data for the direct measure of writing ability. The following two ability ratings were summed to produce the direct measure of Written Communication:
(1) Accuracy in Description and (2) Accuracy in Statement Content. The following six ability ratings were summed to produce the direct measure of Knowledge of Grammar, Spelling, and Punctuation: (1) Logical Flow (2) Vocabulary (3) Spelling and Capitalization (4) Punctuation (5) Grammar, and (6) Logical Sentence Structure. All eight ratings were summed to produce a direct Writing Exercise composite score. The Written Communication was classified as a higher-order writing ability, and the Knowledge of Grammar, Spelling and Punctuation was classified as a lower-order writing ability.

Composite Writing Ability Score

For purposes of this study, a Total Writing Ability score was computed by combining the Multiple Choice Total and the Writing Exercise Total. Both the Multiple Choice Total score and the Writing Exercise Total score were converted to standardized scores before they were summed to produce the Total Writing Ability score.

Training Performance Scores

Three training performance measures represented the criterion data used in this study: (1) Academy Graduation, (2) Overall Academy Grade, and (3) individual course grades. Academy Graduation included three categories of recruits: those that graduated from the academy, those that failed to graduate due to academic failure, and those that did not graduate due to non-academic reasons (i.e, resignations for personal reasons, health problems, behavior problems). As defined by NOPD training academy standards, academic failures included recruits who failed to achieve a minimum score of 70% in any of the final course grades. Overall Academy Grade was the average of the nine courses grades given during the duration of training. Individual course grades represented recruits final grade in each of these nine courses. Criterion data for recruits who did not complete the training academy consisted of only the individual final grade of completed courses; no overall Academy Grade was reported for this group. Because of this, the sample size for the various correlations differed. Table 6 and Table 7 present a summary of the predictor and criteria variables.

Procedures

Applicants demographic information and test scores on the Written Test and Writing Exercise were collected from files of The New Orleans Civil Service Department of Test Development and Validation. Training performance data for each recruit were also provided by the New Orleans Civil Service Department of Test Development and Validation.

Table 6

Predictor Variables

Predictor	Variable
Written Exercise	(1) Written Communication (WC)
	(2) Form Completion (FC)
	(3) Grammar, Spelling, and Punctuation (GSP)
	(4) Multiple Choice Total (MCT) = (WC+FC+GSP)
Writing Exercise	(5) Accuracy in Description (AD)
	(6) Accuracy in Statement Content (ASC)
	(7) Logical Flow (LF)
	(8) Vocabulary (V)
	(9) Spelling and Capitalization (SC)
	(10) Punctuation (P)
	(11) Grammar (G)
	(12) Logical Sentence Structure (LSS)
	(13) Content & Completeness (CC) = (AD+ASC)
	(14) Writing Mechanics (WM) = (LF+V+S&C+P+G+LSS)
	(15) Writing Exercise Total (WET) = (AD+ASC+LF+V+S&C+P+G+LSS)
Writing Composite	(16) Writing Total (WT) = (MCT+WET)

Table 7

Criteria Variables

Criteria	Variable
Academy Performance	(1) Graduation
	Graduate
	Non-graduate due to academic failure
	Non-graduate due to non academic reason
	(2) Overall Academy Grade - Average of Course Grades
	Individual Course Grades
	(3) Legal Aspects
	(4) First Aid
	(5) Investigations
	(6) Report Writing
	(7) Traffic
	(8) Patrol Activities
	(9) Officer Survival
	(10) Specialized Activities
	(11) Police Community Relations

RESULTS

This section presents information in the following manner: (1) predictor analysis including descriptive statistics, reliability analysis, factor analysis, and multitrait-multimethod matrix analysis, (2) criterion analysis, including descriptive statistics, (3) correlations between predictor and criterion scores, (4) validity comparisons of direct, indirect, and the combination of measures, and (5) comparisons between the validity coefficients of direct and indirect measures of higher and lower-order writing ability.

Predictor Analysis

Police Recruit Written Test - Multiple Choice

Table 8 displays the means, standard deviations, and coefficients alpha for dimension and Multiple Choice Total (MCT) scores for the multiple choice test for the unrestricted population (N = 3,250). As can be seen, the mean of the Written Communication dimension was relatively high which may indicate that this part of the test was too easy. Most of the applicants scored relatively high on this dimension. However, the standard deviation for this dimension indicates that the scores were relatively dispersed among the high mean. The coefficients alpha, which can be regarded as a measure of reliability, for individual dimensions range from 0.32 to 0.38. The small number of items measured in each dimension could be contributing to these relatively low alphas. When the items of all three dimensions (Written Communication; Form Completion; and Grammar, Spelling and Punctuation) are summed to produce the Multiple Choice Total writing ability score, the coefficient alpha increases to 0.54.

Table 8

Descriptive Statistics and Reliability Estimates for the Multiple Choice Police Recruit Written Test (N = 3,250)

		No. of items	*M*	*SD*	Estimated Alpha
1.	Written Communication (WC)	8	6.52	1.30	0.38
2.	Form Completion (FC)	8	4.80	1.24	0.32
3.	Grammar, Spelling, Punctuation (GSP)	6	3.63	1.41	0.37
4.	Multiple Choice Total (MCT)	22	14.95	2.75	0.54

Police Recruit Writing Exercise

Descriptive Statistics. Table 9 displays the range of ratings, means, and standard deviations for sub-scales, scales and total score for the Writing Exercise for the restricted population ($N = 2,670$). Only those applicants that passed the multiple choice Written Test took the Writing Exercise with the exception of the first 206 applicants who completed both the multiple choice Written Test and the Writing Exercise. For the purpose of norm development, these 206 applicants took the Writing Exercise regardless of their score on the multiple choice Written Test. As can be seen, the mean of the Accuracy in Description sub-scale was relatively low compared to other sub-scales. To achieve high ratings on this sub-scale, applicants must include all aspects of an incident (who, what, why, when, where). The fact that most applicants have yet to receive training on incident report writing may be contributing to this low mean. The vocabulary sub-scale had the highest mean with the lowest standard deviation which may indicate low discriminability among applicants. The distributions of the remaining sub-scales were relatively normal with no serious problems.

Reliability. Reliability estimates for the Writing Exercise are presented in Table 10. For each sub-scale, the coefficient alpha was computed using the sum of ratings for that sub-scale. Because two raters rated each sub-scale twice (i.e, one rating per video), the total number of ratings for a sub-scale is four. The coefficient alpha for the Content and Completeness scale was based on eight ratings as two sub-scales were summed to produce this scale. For the Writing Mechanics scale, the alpha was based on 24 ratings as six sub-scales were summed for this scale. Finally, the coefficient alpha for Total Writing

Table 9

Descriptive Statistics for the Police Recruit Writing Exercise (*N = 2,750*)

		Range of Ratings[a]	*M*	*SD*
1.	Accuracy in Description (AD)	4-20	10.11	4.06
2.	Accuracy in Statement Content (ASC)	4-20	12.29	3.46
3.	Logical Flow (LF)	4-20	13.15	3.39
4.	Vocabulary (V)	4-20	14.14	3.13
5.	Spelling and Capitalization (SC)	4-20	13.18	4.03
6.	Punctuation (P)	4-20	12.39	3.71
7.	Grammar (G)	4-20	13.46	4.12
8.	Logical Sentence Structure (LSS)	4-20	12.86	3.84
9.	Content and Completeness (CC)	8-40	22.39	6.82
10.	Writing Mechanics (WC)	24-120	79.17	17.63
11.	Writing Exercise Total (WET)	32-160	101.57	22.31

[a]For sub-scales, four ratings were summed to produce a total score; the median for these distribution is 12. The median for Content and Completeness is 24. The median for Writing Mechanics is 72. The median for Writing Exercise Total is 96.

Table 10

Reliability Estimates for the Police Recruit Writing Exercise (N = 2,670)

		Coefficient Alpha	Parallel-Form Correlation[a]
(1)	Accuracy in Description (AD)	0.83	0.39
(2)	Accuracy in Statement Content (ASC)	0.85	0.48
(3)	Logical Flow (LF)	0.88	0.61
(4)	Vocabulary (V)	0.85	0.54
(5)	Spelling and Capitalization (SC)	0.89	0.58
(6)	Punctuation (P)	0.88	0.61
(7)	Grammar (G)	0.89	0.63
(8)	Logical Sentence Structure (LSS)	0.89	0.64
(9)	Content and Completeness (CC)	0.88	0.51
(10)	Writing Mechanics (WM)	0.95	0.80
(11)	Writing Exercise Total (WET)	0.95	0.80

[a]Parallel-form reliability coefficients represent correlations between video one (Domestic Disturbance) and video two (Car Theft).

Exercise (WET) was based on 32 ratings as WET was the sum of all eight sub-scales. As the table shows, the coefficients were high for all sub-scales of the test. Parallel forms correlation coefficients were also calculated for the Writing Exercise (Domestic Abuse video sub-scales correlated with Car Theft video sub-scales). The results are also presented in Table 10. The coefficient for the Writing Exercise Total score (0.80) indicates that overall, the two versions of the exam are very similar.

Factor Analysis. In order to test the proposed two scale structure of the Writing Exercise, sub-scale ratings were factor analyzed using the Principal Components Method and Varimax rotation. Because the test was designed to measure two distinct and separate scales (1) Accuracy and Completeness, and (2) Writing Mechanics, two factors were extracted. Results of the factor analysis are presented in Table 11. After rotation, two factors emerged. All six of the proposed Writing Mechanics sub-scales (Punctuation, Grammar, Logical Sentence Structure, Spelling and Capitalization, Vocabulary, and Logical Flow) loaded on factor one. The two proposed Accuracy and Completeness sub-scales (Accuracy in Description and Accuracy in Statement Content) loaded on factor two. With the exception of the Vocabulary and Logical Flow sub-scales, which loaded relatively high on both factors, the results of the factor analysis support the proposed two scale design.

Multitrait- Multimethod Matrix (Campbell & Fiske, 1959). The multitrait-multimethod matrix was presented by Campbell and Fiske as a method to examine convergent and discriminant validity. This analysis requires that at least two different traits be measured by two different methods. For this study, Written Communication (WC) and Grammar,

Table 11

Factor Analysis of the Police Recruit Writing Exercise (N = 2,670)

Sub-scale	Factor 1	Factor 2
Punctuation	0.83	
Grammar	0.80	0.27
Logical Sentence Structure	0.77	0.38
Spelling and Capitalization	0.75	0.29
Vocabulary	0.57	0.48
Accuracy in Description		0.86
Accuracy in Statement Content	0.28	0.84
Logical Flow	0.52	0.63
Initial Eigenvalue	4.53	0.99
Rotation Sum of Squared Loadings	3.18	2.35
Percent of Variance Explained	39.69	29.31
Cumulative % of Variance Explained	39.69	69.00

Note. Extraction method - Principal Components Analysis with Varimax rotation, 2 factors extracted.

Spelling, and Punctuation (GSP) were measured in both the multiple choice Written Test and the Writing Exercise. Table 12 presents the resulting multitrait-multimethod matrix. Results indicate that the convergent validity for WC is low. Also, the discriminant validity is questioned because WC correlated higher with a different trait (GPS) than with a different measure of WC. This suggests that the Written Test and the Writing Exercise may be measuring different Written Communication constructs. Results for GSP are a bit more positive. There is evidence of both convergent and discriminant validity for this trait. Finally, results indicate a relatively strong method variance for the Writing Exercise.

Criterion Analysis

Academy graduation, overall academy grade, and individual academy course grades were used as the criteria data in this study. Because recruits left the academy at various times covered by the study, the sample sizes for each criterion are not equal. Of the 567 recruits who began one of eleven academy classes included in this study, 469 completed training and graduated from the academy, 51 failed to graduate due to poor academic performance, and 47 did not complete the training program for reasons other than poor academic performance. The Graduation variable includes only those that either graduated from or failed the training academy (n = 520). The Overall Academy Grade data includes only those recruits who successfully completed the academy. Unfortunately, because two academy classes were missing several course grades, Overall Academy Grade data for 106 recruits from these two classes were not included in Overall Academy Grade analysis. It was felt that the Overall Academy Grade for these two classes was not

Table 12

Multitrait- Multimethod Matrix (Campbell & Fiske, 1959) for the Police Recruit Written Test and Writing Exercise

		Multiple Choice		Writing Exercise	
		A_1	B_1	A_2	B_2
Written Test - Multiple Choice					
Written Communication	A_1	(0.38)[a]			
Grammar, Spelling & Punctuation	B_1	0.08[a]	(0.37)[a]		
Writing Exercise					
Written Communication	A_2	0.12	0.26	(0.88)	
Grammar, Spelling & Punctuation	B_2	0.14	0.36	0.59	(0.95)

Note. Correlations in parentheses represent coefficient alpha reliability estimates.

[a]Statisitic derived from sample of applicants who completed the Police Recruit Written Test (*N*=3,250). All other statistics are derived from the sample of applicants that completed both the Written Test and the Writing Exercise (*N*=2,670).

equivalent to the Overall Grade of the other academy classes. However, individual course grades reported for these two classes were included in analysis involving individual course grades. Finally, sample sizes vary for individual courses as recruits failed or resigned from the academy at various stages of training. Table 13 presents the various sample sizes of criterion variables.

The means, standard deviations, and minimum and maximum scores for the criterion variables are also presented in Table 13. As can be seen in the table, individual course means increase as sample size decreases. Legal Aspects, Investigations, and Report Writing are considered by police academy staff to be the most difficult courses. Because of this, these classes are taught first so that poor performers are identified and dismissed during the initial stages of training. This decrease in sample size coupled with the corresponding increase in course means may reflect this process; as poor performers are dropped from the academy, course grades go up. Also, these three courses have the largest ranges and standard deviations indicating that they are differentiating performance better than other academy courses. The First Aid and Community Relations courses have high means with relatively low standard deviations. This may indicate that these courses may be too easy to differentiate between good and poor performers.

Correlations Between Predictors and Criteria

Table 14 shows the correlations between training performance measures and dimension and total scores of the multiple choice Written Test. The test demonstrated predictive validity for the training performance measures. The Multiple Choice Total (MCT) score correlated with all but two training variables at the $p < 0.01$ level. The

Table 13

*Descriptive Statistics for Criterion Variable*s

Variable	*n*	*M*	*SD*	Minimum	Maximum
Graduation[a]	520	0.90	0 .30	1.0	0.0
Overall Academy Grade	363	85.83	3.43	74.22	94.12
Legal Aspects	527	81.49	7.10	53.70	96.22
Investigations	500	81.43	5.94	63.0	95.8
Report Writing	451	83.44	8.67	53.4	102.0
Traffic	429	83.41	6.49	64.0	99.0
Patrol Activities	413	88.85	5.38	69.0	99.00
Officer Survival	319	84.44	6.98	70.0	100.0
Specialized Activities	332	86.96	5.60	70.0	100.0
Community Relations	257	90.77	5.20	75.0	100.0
First Aid	384	92.26	4.12	78.67	100.0

[a] Graduation was coded as follows: academic failure = 0 and graduated = 1.

Table 14

Correlation Coefficients Between Dimension and Total Scores on the Multiple Choice Written Test and Academy Performance

		Written Test			
Academy Performance	*n*	WC	FC	GSP	MCT
Graduation	520	0.08	0.08	0.22**	0.21**
Overall Academy Grade	363	0.09	0.21**	0.30**	0.32**
Legal Aspects	527	0.13**	0.21**	0.34**	0.37**
Investigations	500	0.10*	0.13**	0.20**	0.23**
Report Writing	451	0.08	0.16**	0.27**	0.27**
Traffic	429	0.02	0.11*	0.18**	0.16**
Patrol Activities	413	-0.01	0.09	0.13**	0.11*
Officer Survival	319	-0.01	0.06	0.18**	0.13*
Specialized Activities	332	0.10	0.12*	0.08	0.15**
Community Relations	257	0.17**	0.04	0.24**	0.25**
First Aid	384	0.13**	0.12**	0.30**	0.30**

Note. WC = Written Communication; FC = Form Completion; GSP = Grammar, Spelling, and Punctuation; MCT = Multiple Choice Total.

*$p < 0.05$. **$p < 0.01$.

correlation between MCT and Overall score, which is a measure of overall academy performance, was 0.32. The highest correlation (0.37) was found between MCT and Legal Aspects. The smaller correlations were found between the predictor and Patrol Activities and Officer Survival. This supports the discriminant validity of the Written Test as these two academy courses teach practical skills such as radio procedures, baton handling, and weapon retention.

Table 15 presents correlations between training performance measures and sub-scale, scale, and total scores on the Writing Exercise. The test demonstrated predictive validity of training performance measures that require writing skills. The Writing Exercise Total (WET) correlated significantly with seven of the eleven training variables at the $p < 0.01$ level. The correlation between WET and Overall Academy Grade was 0.24. The highest correlation (0.30) was found between WET and Report Writing. Non-significant correlations were found between the predictor and four academy classes: Traffic, Patrol Activities, Officer Survival, and Specialized Activities. This demonstrates discriminant validity of the Writing Exercise as success in these courses is not, for the most part, based on writing ability.

Validity Comparisons of Direct, Indirect, and the Combination of Measures

Table 16 presents the comparisons of validity coefficients between the multiple choice Written test (an indirect measure), the Writing Exercise (a direct measure), and Writing Total which is the combined sum of both measures. Based on the literature review, it was hypothesized that (1) an indirect measure, the Written Test, would correlate the lowest with academy performance and, (2) the combination of indirect and

Table 15

Correlation Coefficients Between Sub-scales, Scales and Total Scores on the Writing Exercise and Academy Performance

		Writing Exercise										
Performance	*n*	AD	ASC	LF	V	SC	P	G	LSS	CC	WM	WET
Graduation	520	0.17**	0.16**	0.12**	0.08	0.10*	0.15**	0.15**	0.14**	0.19**	0.17**	0.21**
Overall	363	0.13*	0.11*	0.11*	0.06	0.24**	0.25**	0.20**	0.16**	0.14**	0.24**	0.24**
Legal Aspects	527	0.17**	0.16**	0.12**	0.07	0.13**	0. 18**	0.23**	0.18**	0.18**	0.21**	0.24**
Investigations	500	0.22**	0.17**	0.16**	0.16**	0.07	0.15**	0.09*	0.18**	0.22**	0.18**	0.22**
Report Writing	451	0.22**	0.16**	0.17**	0.14**	0.28**	0.23**	0.17**	0.18**	0.22**	0.27**	0.30**
Traffic	429	-0.03	-0.01	-0.01	-0.06	0.10*	0.13**	0.12*	0.09	-0.02	0.09	0.07
Patrol Activities	413	0.08	0.03	-0.03	0.02	0.08	0.12*	0.13**	0.06	0.06	0.09	0.09
Officer Survival	319	0.02	0.03	-0.05	-0.12*	0.07	0.09	0.06	0.07	0.03	0.04	0.04
Special Activities	332	0.05	0.03	0.00	0.01	0.06	0.13*	0.07	0.06	0.04	0.08	0.08
Community Relations	257	0.15*	0.08	0.14*	0.11	0.19**	0.11	0.12*	0.13 *	0.13*	0.19**	0.20**
First Aid	384	0.18**	0.09	0.14**	0.15**	0.17**	0.18**	0.19**	0.23**	0.16**	0.24**	0.25**

Note. AD = Accuracy in Description; ASC = Accuracy in Statement Content; LF = Logical Flow; V = Vocabulary; SC = Spelling and Capitalization; P = Punctuation; G = Grammar; LSS = Logical Sentence Structure; CC = Content and Completeness; WM = Writing Mechanics; WET = Writing Exercise Total.

*$p < 0.05$. **$p < 0.01$.

Table 16

Comparisons of Validity Coefficients Between Indirect, Direct, and Combined Measures

Academy Performance	*n*	Indirect (MCT)	Direct (WET)	Writing Total (WT)
Graduation	520	0.21**	0.21**	0.25**
Overall Grade	363	0.32**	0.24**	0.36**
Legal Aspects	527	0.37**	0.24**	0.36**
Investigations	500	0.23**	0.22**	0.28**
Report Writing	451	0.27**	0.30**	0.35**
Traffic	429	0.16**	0.07	0.14**
Patrol Activities	413	0.11*	0.09	0.13**
Officer Survival	319	0.13*	0.04	0.10
Specialized Activities	332	0.15**	0.08	0.15**
Community Relations	257	0.25**	0.20**	0.28**
First Aid	384	0.30**	0.25**	0.34**

Note. MCT = Multiple Choice Total; WET = Writing Exercise Total.

*$p < 0.05$. **$p < 0.01$.

direct measures (Writing Total) would correlate the highest with academy performance. As seen in Table 16, this hypothesis was only partially supported. The combination of the Written Test and the Writing Exercise had the highest correlation with most academy measures. However, the indirect measure (Written Test) correlated higher than the direct measure (Writing Exercise) on all academy performance measures with the exception of Report Writing. However, the differences between these correlations were small, ranging from -0.03 to 0.09. Overall, these results indicate that the most valid report writing ability measure includes both a direct and indirect measure.

Validity Comparisons Between Direct and Indirect Measures of Higher-order and Lower-order Writing Ability

Based on reviewed research, it was hypothesized that (1) higher-order writing skills are more reliably assessed with direct measures, and (2) lower-order writing skills are more reliably assessed with indirect measures. Because both the multiple choice Written Test (an indirect measure) and the Writing Exercise (a direct measure) contain measures of higher-order and lower-order writing abilities, this hypothesis was tested by comparing the indirect measure to the direct measure for each type of writing order ability (higher vs. lower). Table 17 presents comparisons of the validity coefficients between direct and indirect measures for higher-order and lower-order writing ability. As seen in the table, for higher-order writing ability, direct measure (Writing Exercise) correlations were higher than indirect measure (multiple choice Written Test) correlations across all but three of the academy performance measures. For lower-order writing ability, the validity coefficients of the indirect measure were higher than or equal to the validity

Table 17

Comparisons of Validity Coefficients Between Direct and Indirect Measures of Higher and Lower-order Writing Ability

		Higher Writing Ability		Lower Writing Ability	
Academy Performance	*n*	Indirect	Direct	Indirect	Direct
Graduation	520	0.08	0.19**	0.22**	0.17**
Overall Academy Grade	363	0.09	0.14**	0.30**	0.24**
Legal Aspects	527	0.13**	0.18**	0.34**	0.21**
Investigations	500	0.10*	0.22**	0.20**	0.18**
Report Writing	451	0.08	0.22**	0.27**	0.27**
Traffic	429	0.02	-0.02	0.18**	0.09
Patrol Activities	413	-0.01	0.06	0.13**	0.09
Officer Survival	319	-0.01	0.03	0.18**	0.04
Specialized Activities	332	0.10	0.04	0.08	0.08
Community Relations	257	0.17**	0.13*	0.24**	0.19**
First Aid	384	0.13**	0.16**	0.30**	0.24**

*$p < 0.05$. **$p < 0.01$.

coefficients of the direct measure. These results support the hypothesis that higher-order writing skills such as idea generation are measured best with essay type tests while lower-order writing skills such as grammar, spelling, and punctuation, are assessed best with multiple choice measures.

DISCUSSION

The purpose of this study was to evaluate the effectiveness of the Police Recruit Writing Exercise which is a direct test of both higher-order and lower-order writing abilities. This test was designed to address report writing deficiencies experienced by police in the training academy. Descriptive statistics were examined, and relationships between the Writing Exercise and the writing ability dimensions included on the multiple choice Police Recruit Written Test, an indirect measure of writing skills, were investigated. For criterion validation purposes, Writing Exercise scores and multiple choice Written Test scores were correlated with several training academy performance measures, and the predictive validities of both tests were compared. Finally, because the multiple choice Written Test and the Writing Exercise were designed to assess both higher-order and lower-order writing abilities, comparisons were made to determine which type of test (direct vs. indirect) is most appropriate for assessing different types of writing skills (higher-order vs. lower-order).

Overall, the descriptive statistics of the Writing Exercise indicate few problems with sub-scales being too easy or difficult. However, the vocabulary sub-scale mean was relatively high with a low standard deviation. This indicates that this sub-scale may be too easy and is not distinguishing between good and poor performance. Also, the Accuracy in Description sub-scale had a low mean with a high standard deviation indicating low and highly varied performance on this measure. High ratings on this sub-scale require all aspects of an incident to be reported in complete detail. Lack of applicants' training or

experience in report writing may be contributing to this low mean. The distributions of scores for all other sub-scales were relatively normal, indicting no problems with scale difficulty.

The two scale design of the Writing Exercise was confirmed using Principal Components factor analysis with Varimax rotation. Four of the six sub-scales designed to measure the Grammar, Spelling, and Punctuation KSA (Punctuation, Grammar, Logical Sentence Structure, and Spelling & Capitalization) loaded highly on factor one while the two sub-scales designed to measure Written Communication (Accuracy in Description and Accuracy in Statement Content) loaded highly on factor two. The remaining two sub-scales, Vocabulary and Logical Flow, had moderate loadings on both factors. Analysis of the descriptive statistics suggests that the rating scales may be too easy or ambiguous as the means were relatively high and the standard deviations were relatively low, especially for the Vocabulary sub-scale. A review of the behavioral anchors for these scales suggests that, although defined, there is much room for rater subjectivity. For example, the behavioral anchors for the Vocabulary sub-scale are defined as follows:

> 5 - Complete absence of less acceptable synonyms, slang or words used incorrectly.
>
> 3 - One or more words may be used incorrectly, but they do not change the meaning....
>
> 1 - One or more words are used incorrectly; they substantially change the meaning....

Incorrectly, less acceptable, substantially, and slang are never defined. The Logical Flow rating scale has similar problems with ambiguity.

Test reliability was estimated by coefficients alpha and parallel forms. The coefficient for the Writing Exercise were high for sub-scale, scale and total score, ranging from 0.83 to 0.95. However, the parallel-form reliability estimates were lower, ranging from 0.39 to 0.64 on the sub-scales with the Accuracy in Description sub-scale having the lowest reliability estimate (0.39). This low parallel-form reliability estimate for Accuracy in Description, coupled with a low mean, may indicate that the details of one of the videos may be more complicated than the other. However, the reliability estimate for the Total test of 0.80 indicates that overall, the two versions of the exam are very similar.

Several problems were noted for the multiple choice Written Test. The Written Communication dimension mean was high, indicating that this part of the exam may have been too easy. Also, the dimension reliability estimates were low, ranging from 0.32 to 0.38. The low number of items per dimension may be a contributing factor. When the dimensions were added together, the reliability estimate for the total increased to 0.54.

The multitrait-multimethod matrix analysis indicated that the Written Communication scale lacks convergent and discriminant validity. First, the correlation between the multiple choice measure of Written Communication and the Writing Exercise measure of Written Communication was only 0.12. Second the correlation between the multiple choice measure of Written Communication and the Writing Exercise measure of Grammar, Spelling, and Punctuation was larger than 0.12 as was the correlation between the Writing Exercise measure of Written Communication and the multiple choice measure of Grammar, Spelling, and Punctuation. These results suggest that the Written Communication dimension of each test is not measuring the same trait.

This trait difference could be contributing to the opposing difficulty levels between the two tests on the Written Communication dimension (i.e., the multiple choice Written Communication mean was high whereas the mean of one of the two Writing Exercise Written Communication sub-scales was low). Also, this difference could explain the discrepancy between the multiple choice Written Communication validity coefficient of 0.08 and the Writing Exercise Written Communication significant correlation of 0.19 for predicting Academy Graduation.

Data analysis showed evidence for both convergent and discriminant validity for the Grammar, Spelling, and Punctuation (GSP) dimension. Correlations between different measures of this trait were higher than the correlations between different traits measured by the same method. Also, the validity value for GSP was higher than the correlation obtained between GSP and other variables having neither trait nor method in common. In addition, the parallel-from reliability estimate for the measure of GSP was 0.80, suggesting that the two forms of the test are measuring the same construct.

Research on police selection measures indicated that cognitive tests are valid predictors of academy performance. In their validity generalization study of cognitive measures used in police selection, Hirsh, Northrop, and Schmidt (1986) reviewed 40 studies that used training success as criteria. The mean observed (uncorrected) validity coefficient of the138 validity coefficients presented in the study was 0.36. Ten years later, Aamodt (1997) conducted another meta-analysis of cognitive ability and police performance. Based on the results of 47 samples from 37 studies, Aamodt reported uncorrected validity coefficients of 0.34 for academy grades and 0.22 for academy

graduation. Because both the Police Recruit Written multiple choice Test and the Writing Exercise are measure of cognitive ability, it was predicted that the correlations between these measures and academy performance would be similar to those found in the reviewed studies.

The Writing Exercise demonstrated predictive validity of training performance. Writing Exercise Total significantly correlated with Academy Graduation (0.21) and Overall Academy Grade (0.24). For Academy Graduation, results replicated Aamodt s (1997). For individual course grades, significant correlations, ranging from 0.20 to 0.30, were found between Writing Exercise Total and five of the nine academy courses. Report Writing had the highest correlation with the Writing Exercise (0.30). Evidence of discriminant validity is seen in the low correlations between the Writing Exercise and academy courses such as Patrol Activities, Officer Survival, and Specialized Activities. Designed to teach practical skills such as radio procedures, baton handling, and weapon retention, success in these courses is not, for the most part, based on writing ability. These results indicate that the test is fulfilling the purposes for which is was designed. That is, the Writing Exercise has the highest predictive validity for Report Writing.

The multiple choice Written Test also demonstrated predictive validity of training performance. Multiple Choice Total (MCT) significantly correlated with Academy Graduation (0.21) and Overall Academy Grade (0.32). As was the case for the Writing Exercise, the coefficient for Academy Graduation replicated the results of Aamodt (1997). For academy grades, the validity coefficient of 0.32 was very similar to the coefficients reported by Aamodt (1997) and Hirsh et al. (1986). For individual course

grades, significant correlations, ranging from 0.15 to 0.37, were found between the Multiple Choice Total and all nine academy courses. Legal Aspects had the highest correlation with the multiple choice Written test (0.37). These results further support the use test of cognitive abilities for the selection of police recruit applicants.

Research on written communication suggests that indirect and direct measurements assess different writing abilities. The research indicated that the correlations between the two measures are often only moderately correlated. According to Qeullmalz, Capell, and Chou (1982), correlations between the two measures range from 0.43 to 0.68. Ackerman and Smith (1988) reported similar correlations ranging from 0.30 to 0.60. Ackerman and Smith proposed a model of the writing process which consists of three components: planning, translating, and reviewing. According to this model, direct measures of written communication include all three components whereas indirect measures only include the reviewing component. Ackerman and Smith s research generally supported this model. However, they did find that essay scores were almost totally dominated by the higher-order generation components whereas multiple choice measures represented editing and reviewing skills. Based on these findings, the researchers suggested that the most reliable predictors of writing performance would include both a direct and indirect measure.

Based on this research, the correlation between the Writing Exercise and the multiple choice Written Test should fall in the range of 0.30 to 0.68. Also, it was hypothesized that the combination of the multiple choice Written Test score and the Writing Exercise score would be a better predictor of academy performance than either

measure by itself. Finally, because report writing mainly involves many higher-ordered writing skills such as idea generation and paragraph developments, it was hypothesized that the Writing Exercise would be a better predictor of training performance than the multiple choice Written Test.

The correlation of 0.40 ($p < 0.01$) between the Writing exercise and the multiple choice Written Test measure was similar to correlations found in the reviewed literature. However, only one of the hypotheses concerning direct verses indirect measures was supported. Of the three predictor measures (Writing Exercise Total score, Multiple Choice Total score, and composite score), the composite score had the highest correlation with Overall Academy Grade (0.36, $p < 0.01$) and Academy Graduation (0.25, $p < 0.01$). The hypothesis that the Writing exercise would correlate higher than the multiple choice Written Test measure with academy performance was not supported. For Academy graduation, the correlations of the two predictor tests were equal (0.21, $p <$0. 01). For Overall Academy Grade, the Multiple Choice Total score correlation was 0.32 whereas the Writing Exercise Total score was 0.24. For individual course grades, the multiple choice Written Test measure had higher correlations than the Writing Exercise on all courses with the exception of Report Writing. However, the differences between the correlations were relatively small, ranging from -0.03 to 0.09. At first, these results appear surprising. However, they suggest that the multiple choice Written Test may be assessing dimensions of writing ability that are needed for overall success in the academy whereas the Writing Exercise is assessing writings skills that are more specific to report writing.

The research of Qeullmalz, Capell, and Chou (1982) and Ackerman Smith (1988) indicated that lower-order writing abilities such as spelling, capitalization, and punctuation are best assessed with an indirect measure. More complex, higher-order writing abilities, such as idea and paragraph development, are best measured with a direct, essay type test. Based on the reviewed research, comparisons between direct and indirect measure of higher-order and lower-order writing ability were made to test the following hypotheses: (1) for higher-level writing abilities, a direct measure would have higher validity coefficients than an indirect measure, and (2) for lower-level writing abilities, an indirect measure would have higher validity coefficients than a direct measure.

The results of the study supported both hypotheses. For the Written Communication Knowledge Skill and Ability (KSA), a measure of higher-order writing ability, the Writing Exercise had higher validity coefficients than the multiple choice test for both Academy Graduation and Overall Academy Grade. This supports hypothesis one. For the Grammar, Spelling, and Punctuation KSA, which was classified as a lower-order writing ability, the multiple choice test had higher validity coefficients than the Writing Exercise for both Academy Graduation and Overall Academy Grade. This supports the second hypothesis.

The reviewed research on written communication only concentrated on differences between the direct and indirect measures; no correlations between the tests and a performance measure were reported. However, in the reviewed literature of validation studies of cognitive assessment in police work, a small number of studies did

correlate a measure of writing ability with performance in the training academy. For direct measures of higher-order writing skills, validity coefficients of 0.23 (DuBois & Watson, 1950) and 0.30 (Berkley, 1997) were reported. For direct measures of lower-order writing ability, validity coefficients ranging from 0.24 to 0.58 were reported (Rafilson & Sison, 1996). For indirect measures of lower-order writing ability, Rafilson and Sison reported validity coefficients ranging from 0.20 - 0.42. None of the reviewed studies included an indirect measure of a higher-order writing ability; thus, no validity coefficients could be reported.

Because the multiple choice Written Test was designed to measure a high-order writing ability (Written Communication), validity coefficients for an indirect measure of a higher-order writing ability were computed. The correlations between the multiple choice Written Communication dimension and Academy Graduation and Academy Grade were both low and non-significant (0.08 and 0.09). However, because the mean for the Written Communication dimension was so high, 6.52 out of a total of eight, this validity coefficient, although low, is not surprising. The lack of variance in the Written Communication scores is definitely influencing the resulting validity coefficient.

Implications

The importance of police report writing skills is well documented in the criminal justice literature (Johnson, 1987; Miller & Pomerenke, 1989; Stanard & Associates, Inc.,1992; Wilson & Hayes, 1984). However, the law enforcement community has been complaining about the poor quality of reports for years (D Aulizio & Sheehan; 1992b). One of the biggest reasons for these report problems can be attributed to the fact that

many recruits do not possess the necessary writing skills needed to complete a police report (Wilson & Hayes, 1984; D Aulizio & Sheehan, 1992b). Thus, it is essential for police departments to utilize selection measures that accurately identify individuals who lack the skills needed for successful academy training (Boehm, Honey, & Kohls, 1983). Unfortunately, few empirical studies have explored the assessment of written communication for personnel selection. This study attempted to integrate the findings of previous work in the areas of personnel selection and written communication in order to validate an instrument designed to assess report writing skills. Overall, the validity of the Police Recruit Writing Exercise as a measure of report writing ability was supported.

In addition, this study also replicated and expanded on educational research on direct and indirect assessment of written communication. Results confirmed previous research findings that support the use of direct, essay type tests to assess higher-order writing ability and indirect, multiple choice measures to assess lower-order writing skills. This study took the educational research one step forward by applying the theoretical framework of written communication to workplace performance data. Finally, the validity of an indirect measure of higher-order writing ability was investigated, which is an area that has not been addressed. Unfortunately, results were un-interpretable due to low variance in the predictor measure.

Limitations

Overall, both the multiple choice Written Test and the Writing Exercise appear to be valid predictive measures of academy performance. However, in-depth analysis suggests that the validity coefficients were reduced as a result of problems with various

sub-components of the tests. The major limitation that appeared throughout this study is related to problems with the Written Communication (WC) construct. High means were reported for the multiple choice WC measure whereas the means for the Writing Exercise WC measure were relatively low. Also, and possibly related to the lack of variance in the multiple choice measure, results of the multitrait-multimethod analysis question the convergent and discriminant validity of the Written Communication measure. On one hand, the results suggest that the multiple choice WC dimension is not measuring the same abilities as the Writing Exercise measure. However, it may be that the problem is with the test itself, not with the measurement of different abilities. The analysis of an indirect measure of higher-order writing skills was also limited by problems with the Written Communication measure. The resulting validity coefficients were difficult to interpret due to the low variance in the Written Communication measure.

There were also problems with the Writing Exercise Vocabulary and Logical Flow sub-scales. Although these sub-scales were designed to measure the Grammar, Spelling, and Punctuation KSA, they had moderate loadings on both the Written Communication factor and the Writing Mechanics factor. The vocabulary sub-scale also suffered from problems associated with a high mean and low variance. A closer look at the rating scales of both these measures indicated that rater subjectivity may be a problem as scales were vague and poorly defined.

Finally, there were several problems with the criterion data. First, because recruits are dismissed from the academy as soon as they fail a course, criterion data sets for failures was incomplete. This limits the variance in criterion measures as the performance

of failures is not included in all academy course finals. The Overall Academy Grade was also affected by this range restriction. Final course grades were averaged to produce the Overall Academy Grade measure. Because they were missing one or more final course grades, the Overall Academy Grade was not calculated for the 51 recruits coded as academic failures. Also, because the police academy failed to report several course grades for recruits in two academy classes, the Overall Academy Grade was not calculated for the 106 recruits in these classes. Thus, besides the normal range restriction issues associated with criterion variables, this study was further affected by range restriction encountered in the academy. Finally, two academy courses had high means and relatively low standard deviations. Because these courses are included in the calculation of Overall Academy Grade, their reduced variance may be lowering the validity coefficient.

Future Research

Overall, the Writing Exercise appeared to have adequate predictive validity for academy performance. However, some tweaking of the scales may increase the validity coefficients. Results indicated that the rating scales for Vocabulary and Logical Flow may need to be reworded for better clarity and increased difficulty. Also, further research needs to be done on the written communication construct as problems with this scale contributed to difficulty in interpreting several analysis in this study. It is recommended that the test items for the multiple choice measure of Written Communication should be made more difficult to increase the variance in that portion of the exam. Finally, as a measure of test stability, research should be conducted to determine if sub-scale written communication scores increase as recruits gain more field experience in report writing.

Further research in personnel selection is needed on the direct assessment of writing skills for entry workers. While direct assessment appears to provide more information on global writing skills, practical issues associated with the time and expense of rating these types of exams warrants careful consideration. Because different jobs require different writing abilities, direct measure of writing ability may not always be the most appropriate.

APPENDIX

WRITTEN SUB-TESTS EXAMPLE TEST QUESTIONS

Written Sub-tests Example Test Questions

Written Communication--Examinees choose the alternative which is the most accurate and clearly written description of a corresponding picture.

Example:

E1. Choose the most accurate and clearly written description of picture P.

A. The blood and glass inside the car could indicate that the victim was shot by someone outside the car.

B. There was blood and bullet holes and a big mess all over the car seat.

C. The seat of the car was a mess, especially because of the blood.

D. The glove compartment was open and the car was a mess.

Form Completion--Examinees are given instructions regarding how to fill out an Event section of an Incident Report and then provided with completed reports in which they determine if there are any mistakes or errors.

Example:

E2. An officer in the course of crowd control duties inside the Superdome (Sugar Bowl Dr.) notices a man carrying a concealed gun. The barrel of the weapon was falling through a hole in the suspect s back pocket. This happened at 1:00 P.M., 8-1-98. Another officer arrived at the scene 15 minutes later to take the report. The Superdome is located in the 1st District. Outside it was raining and 82 degrees Fahrenheit. The lighting was good.

NOPD INCIDENT REPORT		
EVENT SECTION		
1. Signal: 95-G	2. Incident: Carrying concealed gun	3. Date/Time of Occurrence: 8-1-98
4. District: 1st	5. Location of Occurrence: Superdome	6. Date/Time of Report: 8-1-98, 1:15 PM
7. Weather: Rain	8. Temperature: 82 degrees	9. Lighting: Good

Which box is filled out INCORRECTLY:

(1) Box #1
(2) Box #2
(3) Box #3
(4) Box #7
(5) none of the above

Grammar, Spelling, and Punctuation--Examinees find the sentence without grammar, spelling, or punctuation mistakes.

E3. Please read each sentence for grammar, punctuation, and spelling and select the one that is most correct.

A. Officer Smith and Jones returned to Officer Jones car to radio for assistance.

B. Officers Smith and Jones returned to his car to radio for assistance.

C. Officer's Smith and Jones returned to Jone's car to radio for assistance.

D. Officers Smith and Jones returned to Officer Jones' car to radio for assistance.

REFERENCES

Aamodt, M. G. (1997, April). *Relationship between cognitive ability and police performance: a meta-analysis.* Paper presented at the annual meeting of the Society for Industrial and Organizational Psychology, St. Louis, MO.

Ackerman, T. A., & Smith, P. L. (1988). A comparison of the information provided by essay, multiple choice, and free response writing tests. *Applied Psychological Measurement, 12(2),* 117-128.

Berkley, S. (1997, June). *Selecting Corrections Officers.* Paper presented at the annual meeting of the International Personnel Management Association Assessment Council, Newport Beach, CA..

Boehm, N. C., Honey, R., & Kohls, J. (1983). Predicting success in academy training: the POST reading and writing test battery. *The Police Chief, 50(10),* 28-31.

Breland, H. M., & Gaynor, J. L. (1979). A comparison of direct and indirect assessments of writing skills. *Journal of Educational Measurement, 16(2),* 119-128.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

D Aulizio, M. T., & Sheehan, K. M. (1992a). Writing reports in an age of lawsuits. *Law and Order, 40(3),* 112-114.

D Aulizio, M. T., & Sheehan, K. M. (1992b). Instituting quality control measures for police reports. *The Police Chief, 59,* 129-132.

DuBois, P. H., & Watson, R. I. (1950). The selection of patrolmen. *Journal of Applied Psychology, 34,* 90-95.

Ford, J. K., & Kraiger, K. (1993). Police officer selection validation project: the multijurisdictional police officer examination. *Journal of Business and Psychology, 7(4),* 421-429.

Godwin, P. A. (1993). Painless report writing. *Law and Order, 41(2),* 38-40.

Gruber, G. (1986). The police applicant test: a predictive validity study. *Journal of Police Science and Administration, 14(2),* 121-129.

Hess, K. M., & Wrobleski, H. M. (1991). *For the record: Report writing in law enforcement.* Blue Lake, CA: Innovative Systems.

Hirsh, H. R., Northrop, L. C., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology, 39,* 399-420.

Johnson, D. M. (1987). Attacking the report writing problem. *The Police Chief, 54,* 22-26.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult Literacy in America: a First Look at the Findings of the National Adult Literary Survey* [On-line]. Available: www.nces.ed.gov/naal/naal92/Overview.html

Kleiman, L. S., & Gordon, M. E. (1986) An examination of the relationship between training academy performance and job performance. *Journal of Police Science and Administration, 14(4),* 293-299.

McAllister, J. A. (1970). A study of the prediction and measurement of police performance. *Police, 14,* 58-64.

Miller, M., & Pomerenke P. (1989). Police reports must be reader based. *Law and Order, 37(9),* 66-69.

Moss, P. A., Cole, N. S., & Khampalikit, C. (1982). A comparison of procedures to assess written language skills at grades 4, 7, and 10. *Journal of Educational Measurement, 19(1),* 37-47.

Pettaway, M. M. (1994). *The development of the New Orleans Police Department Written Communication Test.* Unpublished doctoral dissertation, Tulane University, New Orleans.

Quellmalz, E. S., Capell, F. J., & Chou, C. (1982). Effects of discourse and response mode on the measurement of writing competence. *Journal of Educational Measurement, 19(4),* 241-258.

Rafilson, F., & Sison, R. (1996). Seven criterion-related validity studies conducted with the national police officer selection test. *Psychological Reports, 78(1),* 163-176.

Spielberger, C. D., Spaulding, H. C., Jolly, M. T., & Ward, J. C. (1979). Selection of effective law enforcement officers: The Florida police standards research project. In C. D. Spielberger (Ed.), *Police selection and evaluation: Issues and techniques* (pp.231-251). New York: Praeger Publishers.

Stanard & Associates, Inc., (1991). *The national police officer selection test: Technical information.* Chicago, IL: Author.

Stanard & Associates, Inc. (1992). Poor quality of applicants raises hiring concerns. *Quality Through Selection Newsletter,* Winter, 1-3.

Strobl, W. M. (1984). *The investigator s handbook.* Boston: Butterworth Publishers.

Sulzer, J. L. (1986). An evaluation of the education and training division of the New Orleans Police Department. (Conducted for the Chief Administrative Officer of the City of New Orleans). New Orleans, LA.

Wilson, J. B., & Hayes, S. P. (1984). The competence-based approach to report writing: a new option. *The Police Chief, 51,* 26-29.